

Fuzzy RNA recognition by the *Trypanosoma brucei* editosome

Wolf-Matthias Leeder, Felix Klaus Geyer and Hans Ulrich Göringer¹*

Molecular Genetics, Technical University Darmstadt, 64287 Darmstadt, Germany

Received March 18, 2022; Revised April 20, 2022; Editorial Decision April 20, 2022; Accepted April 26, 2022

ABSTRACT

The assembly of high molecular mass ribonucleo-protein complexes typically relies on the binary interaction of defined RNA sequences or precisely folded RNA motifs with dedicated RNA-binding domains on the protein side. Here we describe a new molecular recognition principle of RNA molecules by a high molecular mass protein complex. By chemically probing the solvent accessibility of mitochondrial pre-mRNAs when bound to the *Trypanosoma brucei* editosome, we identified multiple similar but non-identical RNA motifs as editosome contact sites. However, by treating the different motifs as mathematical graph objects we demonstrate that they fit a consensus 2D-graph consisting of 4 vertices (V) and 3 edges (E) with a Laplacian eigenvalue of 0.5477 (λ_2). We establish that synthetic 4V(3E)-RNAs are sufficient to compete for the editosomal pre-mRNA binding site and that they inhibit RNA editing *in vitro*. Furthermore, we demonstrate that only two topological indices are necessary to predict the binding of any RNA motif to the editosome with a high level of confidence. Our analysis corroborates that the editosome has adapted to the structural multiplicity of the mitochondrial mRNA folding space by recognizing a fuzzy continuum of RNA folds that fit a consensus graph descriptor.

INTRODUCTION

Mitochondrial gene expression in the protozoan parasite *Trypanosoma brucei* relies on a nucleotide-specific RNA-editing reaction. In the process sequence deficient and as a consequence nonfunctional primary transcripts are remodeled to functional messenger (m)RNAs by the site-specific insertion and deletion of uridine (U)-nucleotides (nt) (1). Depending on the transcript the extent of the reaction can vary drastically. While only 4 U-nt are inserted into the cytochrome oxidase II (COII) transcript, more than 600 U-nt are processed in the case of the NADH-dehydrogenase sub-

unit 7 (ND7) mRNA. For nine, so-called pan-edited transcripts >50% of the mature mRNA sequence are the result of the processing reaction, and since edited mRNAs code for key components of the mitochondrial ribosome and the mitochondrial electron transport and chemiosmosis system the RNA-processing reaction is essential for the organism.

The catalytic machinery of the process is a mitochondrial multiprotein complex termed the editosome. The 800 kDa particle provides a catalytic surface for all steps of the reaction cycle including endo- and exonuclease, terminal uridylyl-transferase (TUTase), RNA ligase and RNA-chaperone activities (2). Accessory enzymes contribute as well and involve RNA-annealing (3–5) and RNA-helicase activities (6–8). The sequence specificity for the U-insertion/U-deletion reaction is provided by 40–60nt long, non-coding (nc)RNAs termed guide (g)RNAs (9). Guide RNAs hybridize to the pre- and partially edited mRNAs and control the U-nt insertion/deletion reaction by antiparallel base pairing. At steady-state >1000 different gRNAs are expressed in the mitochondria of insect-stage African trypanosomes (10) and in the majority of cases, multiple (≤ 10) gRNAs are needed to fully edit a single pre-mRNA.

The initial step of the editing reaction is the binding of a pre-edited transcript into the single RNA-binding site of the editosome (11). The reaction is assisted by multiple oligonucleotide/oligosaccharide-binding (OB)-fold proteins on the surface of the editosome, which collectively execute a chaperone-type RNA-unfolding activity (12). The activity is driven by the intrinsically disordered protein (IDP)-domains of the different polypeptides (13) and as a result, the highly folded pre-mRNAs (14) become partially unfolded thereby increasing their probability to interact with a gRNA molecule to initiate the reaction cycle (12,15). Pre-mRNA/editosome complexes can be formed *in vitro* (11) and pre-mRNA mimicking oligoribonucleotides become accurately edited in a gRNA-dependent manner (16). Together, this demonstrates that the pre-mRNA binding reaction is an editosome-inherent trait, which, at least *in vitro*, can be executed in the absence of additional protein factors. Intriguingly, the editosomal RNA-binding capacity is characterized by staggering plasticity. For instance, editosomes process a highly pleomorphic set of RNA ligands. This includes pre-edited mRNAs varying in length from

*To whom correspondence should be addressed. Tel: +49 61511624642; Email: goringer@bio.tu-darmstadt.de

164nt (CR3) to 1117nt (Cyb) and thousands of partially edited mRNAs (17). Short, synthetic oligoribonucleotides of only 30nt are edited as well, as are phosphorothioate- and/or 2'-O-CH₃-modified oligoribonucleotides (16). Even more perplexing, editosomes maintain their RNA-binding capacity even though the pre-mRNAs continuously change their primary sequence by being edited. Pan-edited transcripts double their U-content from 30% to 60% during the processing reaction, thereby altering their R/Y-ratio from R-rich (R/Y = 1.9) to Y-rich (R/Y = 0.6). How editosomes solve this 'moving-target problem' is not understood (Supplementary Figure S1). In the same way, the RNA-binding motif that facilitates binding to the editosome is not known.

As demonstrated for many RNA/protein complexes, complex formation often results in the burial of parts of the RNA surface deep inside the protein contact site (18,19). As a consequence, these RNA sequences become secluded from the surrounding solvent and can be identified by comparing the RNA-solvent accessibility in the free and protein-bound states. Here we determine the changes in solvent accessibility of five pre-edited mRNAs when bound to the *T. brucei* editosome using hydroxyl radical footprinting (HRP) (Figure 1A and Supplementary Figure S2). Hydroxyl radicals cleave the sugar-phosphate backbone of accessible ribonucleotides and with a size of only 0.36 nm, they are perfectly suited for solvent probing experiments (20,21). We performed high-throughput HRP-experiments by identifying HR-cleavage positions as abortive cDNA-synthesis products using fluorophore-derivatized oligodeoxynucleotide primer molecules in combination with capillary electrophoresis (CE) and laser-induced fluorescence (LIF) detection (22,23) (Figure 1A). We identify one or two editosome-dependent RNA footprints in all five pre-mRNAs. The different signatures are similar but not identical, which suggests a malleable, fuzzy-type RNA-recognition mode by the editosome similar to what has been described for fuzzy protein/protein complexes (24). Using graph-theory (25) we characterize the fuzzy recognition motif to consist of four vertices (*V*) and three edges (*E*) and confirm that synthetic 4*V*(3*E*)-RNAs are sufficient to compete for the editosomal pre-mRNA binding site and that they inhibit RNA editing *in vitro*. Furthermore, we investigate the binding of a topological RNA motif library to the editosome and we identify that only two topological descriptors are necessary to predict the RNA binding potential of any RNA motif to the editosome. Our analysis demonstrates that the *T. brucei* editosome is competent to process the structural multiplicity or dynamic disorder in the pre- and partially edited mRNA-folding space by recognizing a fuzzy continuum of RNA-binding motifs. This rationalizes the enigmatic RNA-binding characteristics of the *T. brucei* editosome and provides a simple mechanism on how the high molecular mass protein complex can bind and process the structurally pleomorphic pool of pre- and partially edited mitochondrial mRNAs.

MATERIALS AND METHODS

Oligonucleotide synthesis

DNA and RNA oligonucleotides were synthesized by automated solid-phase synthesis using controlled

pore glass (CPG)-beads and 2-cyanoethyl- (DNA) or 2'-O-(tert-butyl)dimethylsilyl (TBDMS)-protected (RNA) phosphoramidites. Fluorophore modifications (6-carboxy-fluorescein = FAM, 5'-dichlorodimethoxy-fluorescein = JOE, and 5-Carboxytetramethylrhodamine = TAMRA) were introduced post-synthetically using a 5'-terminal C₆-amino linker, which were conjugated by EDC (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide)-mediated coupling. Similarly, 3'-end NH₂-modifications were introduced via a hexamethylene-amino linker. Base-protecting groups were removed using NH₄OH/EtOH (3:1) at RT and 2'-silyl protecting groups were removed using neat triethylamine tri-hydro fluoride. All oligonucleotides were HPLC-purified, analyzed by mass spectrometry, and further scrutinized in denaturing polyacrylamide gels (Supplementary Figure S3). Oligonucleotide concentrations were derived from UV-absorbency measurements at 260 nm (*A*₂₆₀). Oligonucleotide sequences are listed in Supplementary Table S1.

Pre-edited mRNA synthesis

Pre-edited mRNAs (A6, ND7, RPS12, COIII, ND3) were synthesized by run-off *in vitro* transcription from linearized plasmid DNA templates. Typically, reactions were performed in 0.1 ml in 40 mM Tris-HCl pH 7.9, 6 mM MgCl₂, 2 mM spermidine, 10 mM DTT containing 3 μg linearized DNA-template, 1.5 mM of each ATP, CTP, GTP, UTP, 40 U Ribolock RNase inhibitor (Thermo Fischer Scientific) and 100 U T7-RNA polymerase. Samples were incubated for 2 h at 37°C and stopped by DNa-seI digestion (15 min, 37°C) followed by phenol extraction. Non-incorporated ribonucleotides were removed by size exclusion chromatography. RNA concentrations were calculated from UV-absorbency measurements at 260 nm (*A*₂₆₀). The integrity of the different pre-mRNA preparations was analyzed in 6% (w/v) urea-containing (8 M) polyacrylamide gels. Radiolabeled (³²P) pre-mRNA preparations were synthesized in 0.02 ml reactions as detailed above using 1 μg linearized DNA-template, 6 μM α-(³²P)-UTP (400 Ci/mmol), 12 μM UTP and 50 U T7-RNA polymerase.

RNA motif library synthesis

RNA tree graph topologies of 5*V*(4*E*), 4*V*(3*E*), 3*V*(2*E*) and 2*V*(1*E*) were formed by intra- and intermolecular base-pairing of oligoribonucleotides of defined sequences (Supplementary Table S2). The individual RNAs were assembled in 25 mM HEPES-KOH pH 7.8, 30 mM KCl at a final RNA concentration of 88 μM (2 × 44 μM for bimolecular assemblies, 3 × 29.3 μM for trimolecular RNAs, and 4 × 22 μM for tetramolecular RNAs). Samples were heat-denatured at 85°C (2 min) and annealed by cooling to 27°C at a rate of 0.2°C/s. Samples were adjusted to 10 mM MgCl₂ and further thermally equilibrated for 5 min at 27°C. Product formation was analyzed by gel electrophoresis in native or semidenaturing (2 M urea) 18% (w/v) polyacrylamide gels (Supplementary Figure S4). Yields varied between 90–95%.

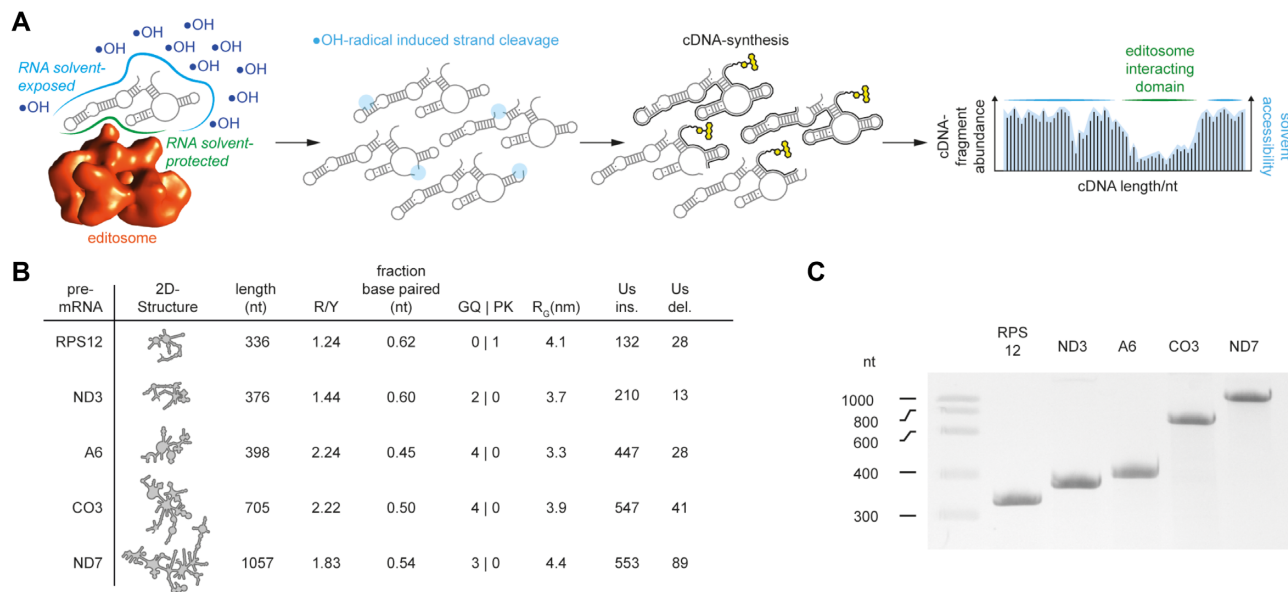


Figure 1. Hydroxyl radical footprinting of editosome/pre-mRNA complexes. (A) 2D-representation of a mitochondrial pre-mRNA bound to the *T. brucei* editosome (red sphere). Complex formation generates an RNA/protein interface (green line), in which parts of the RNA become solvent-protected (green line). Treatment of the complexes with hydroxyl radicals ($\bullet\text{OH}$) results in RNA-backbone cleavage in all solvent-exposed RNA domains (blue line). At single-hit conditions, RNA molecules are cleaved only once (dots in light blue) and strand cleavage positions are mapped by abortive cDNA-synthesis (black lines) using fluorophore-labeled primer molecules (yellow). cDNA fragments are separated by capillary electrophoresis and quantified by laser-induced fluorescence. A plot of the cDNA-fragment abundance in relation to the nt-length represents a read-out of the nt-solvent accessibility. (B) Characteristics of the 5 tested *T. brucei* pre-mRNAs including 2D-structure, nt-length, purine/pyrimidine (R/Y)-ratio, the fraction of base-paired nucleotides (nt), presence of G-quadruplex (GQ)- and pseudoknot (PK)-folds, the radius of gyration (R_G) and the number inserted (ins) and deleted (del) U-residues. (C) Electrophoretic separation of *in vitro* transcribed *T. brucei* pre-mRNAs RPS12, ND3, A6, CO3 and ND7. RNA preparations (3 μg) are separated in a 6% (w/v), urea-containing (8 M) polyacrylamide gel followed by Toluidine Blue O staining. Samples purities are $\geq 97\%$.

Editosome purification

Editosomes were purified from insect-stage African trypanosomes (*Trypanosoma brucei*) grown at 27°C in SDM-79 medium in the presence of 10% (v/v) fetal calf serum (26). The complexes were isolated by tandem affinity purification (TAP) (27) using transgenic *T. brucei* cell lines that conditionally express TAP-tagged versions of the editosomal proteins KREPA3, KREPA4 or KRET2 (28–30). Routinely about 6×10^{11} parasite cells were lysed and further processed using IgG- and calmodulin-affinity chromatography resins. The protein composition of the isolates was analyzed in SDS-containing polyacrylamide gels and the identity of the different proteins was determined by mass spectrometry. The integrity of the complexes was scrutinized by atomic force microscopy (AFM) (11) and the RNA editing activity of the isolates was tested for both the U-insertion and the U-deletion reaction (16,31). The RNA-binding capacity was further quantified by nitrocellulose (NC)-filter binding using (^{32}P)-labeled preparations of pre-edited RPS12-, ND3-, A6-, CO3- and ND7-mRNAs. Reactions were performed in 35 μl EB containing 12 nM pre-mRNA (50 000 cpm) and 15 nM editosomes. Samples were incubated for 30 min at 27°C followed by filtering through NC-filter membranes (Merck Millipore HAWP 0025) at a rate of 7.5 ml/min. NC-filters were washed with 30 reaction volumes (1 ml) EB, dried, and quantified by liquid scintillation counting. Binding activities varied between 87% and 98%.

Hydroxyl radical footprinting (HRP)

In vitro transcribed pre-edited mRNAs (50 nM) were refolded as described in Leeder *et al.* (14). Editosomes in editing buffer (EB = 20 mM HEPES-KOH pH 7.5, 30 mM KCl, 10 mM MgCl_2 , 0.5 mM DTT, 0.5 mM ATP) were added at a molar RNA:editosome ratio of 1:2 and were allowed to bind for 20 min at 27°C. Note: the incubation time is ≥ 5 -times the half-life ($t_{1/2}$) of pre-mRNA/editosome complexes (32) and the RNA concentration is $10 \times K_d$ (11). 27°C represents the optimal temperature for the U-insertion/U-deletion reaction. Control samples contained EB only. Hydroxyl radicals were generated by the sequential addition of 265 μM $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 176 μM EDTA, 0.007% (v/v) H_2O_2 and 4.25 mM Na-L-ascorbate ($\text{C}_6\text{H}_7\text{O}_6\text{Na}$) in a total volume of 35 μl up to 70 μl (22,23,33). Samples were incubated for 45 s at 27°C, which corresponds to single-hit conditions within a 300–400nt sequence window. Reactions were terminated by EtOH precipitation. RNA precipitates were washed (70% (v/v) EtOH), resuspended in 10 mM Tris-HCl pH7.5, 1 mM EDTA (TE)-buffer followed by phenol- CHCl_3 -isoamyl alcohol (25:24:1) extraction and desalting by size exclusion chromatography (Bio-Gel[®] P6-resin). OH-radical induced phosphodiester cleavage sites were identified by abortive reverse transcription using fluorophore-substituted DNA-primer molecules as described in Leeder *et al.* (14). cDNA-fragments were resolved by automated capillary electrophoresis (14,34) and raw electropherograms were pro-

cessed using ShapeFinder (35). HRP-data were statistically analyzed using custom Python scripts with the help of forgi 2.0 (36).

Editosome-binding competition assay

Trimolecular U-deletion pre-mRNA/gRNA hybrid RNAs were formed by combining equimolar concentrations (0.1 μ M) of the pre-mRNA mimicking RNA oligonucleotides 5'Cl22 and 3'Cl15 with guide RNA gRNA_{del}. Samples were denatured for 1 min at 75°C followed by cooling to 27°C at a rate of 0.2°C/s. Annealed pre-mRNA/gRNA-hybrids (100 fmol) in 30 μ l 20 mM HEPES–KOH pH 7.5, 30 mM KCl, 10 mM MgCl₂, 3 mM DTT, 0.5 mM ATP, 60 μ M UTP were mixed with the RNA motif library constructs as competitors for the editosome RNA binding site (0.3- to 5000-fold molar excess) and reactions were started by the addition of editosomes (2.5 nM). After incubation for 30 min at 27°C reactions were stopped by phenol extraction and RNAs were recovered by EtOH-precipitation and analyzed by capillary electrophoresis (CE). Raw electropherograms were baseline corrected followed by peak integration (16,31). Relative RNA-editing efficiencies were calculated as the ratio of edited product RNA over input RNA and were normalized to a control reaction without competitor RNA. Data were plotted as a function of the molar excess of competitor RNA over input RNA and the resulting dose-response curves were fitted to the Hill-Langmuir equation ($\theta_{ed} = 1/(1+(K_d/[RNA])^n)$) to derive half-maximal inhibitory concentrations (θ_{ed} = fraction of RNA-bound editosomes, $[L]/[RNA]$ = free RNA concentration, K_d = RNA concentration at half-maximal occupation, n = Hill-coefficient). A detailed outline of the experiment is shown in Supplementary Figure S7.

RNA-As-Graphs (RAG) analysis

Solvent-protected RNA motifs, defined as 2D-structure elements displaying clustered sequence stretches with negative $\Sigma \Delta$ HRP-values (over a 3nt window) below the 25th percentile were analyzed using the RNA-As-Graphs (RAG) resource developed by Schlick and colleagues (37,38). RAG converts RNA 2D-structures into coarse-grained 2D-tree graphs and classifies them by their vertex (V) number and eigenvalue spectrum.

Calculation of graph topology indices

Graph topology enumerations were performed by calculating numerical indices based on the distance matrix (DM). If G is a connected, unidirectional weighted tree graph with i vertices (V_i) and $i - 1$ edges (E_{i-1}), then the distance matrix of G for any pair of vertices ($V_i V_j$) is defined as the length of the shortest path in bp. From DM, a Wiener index (WI)-type descriptor (39,40) was calculated as $WI = 0.5 \sum e_{ij}$ (e_{ij} = all matrix elements). Graph distance energies (DE) were calculated as $DE = \sum |\lambda_i|_{DM}$ with λ_i specifying the eigenvalues of DM. In addition to DE, we calculated the Frobenius norm of the distance matrix ($\|DM\|_{FN}$) as $\|DM\|_{FN} = \sqrt{(\sum |e_{ij}|^2)}$ (41). Loop topologies were calculated with the help of two descriptors derived

from the dual graph concept (37). If G is a connected, directed and weighted dual graph with the edges (E) representing the loop length in phosphodiester bonds, the adjacency matrix of G is termed loop matrix (LM). Its Frobenius norm ($\|LM\|_{FN}$) is a descriptor of all nt in a loop conformation in analogy to $\|DM\|_{FN}$. The relative connectivity (rC) of the different graphs was defined as the trace of the degree matrix (D) that represents the ratio of the maximal matrix trace ($\text{trace}_{D_{max}}$) over the actual trace (trace_D). For $V = 1$, rC = 0. $\text{Trace}_{D_{max}}$ always corresponds to the linear or circular topology of G with $V \geq 3$, i.e. a multi-loop is $V_i(E_i)$ with $d_{ij} = 2$, while an internal-loop is $V_i(E_j)$ with $j > i$ and $d_{ij} = 2$ or 4. To describe the symmetry aspects of the different RNA constructs, the anisotropy of any internal and/or multiloop sequence (A_{IML}) was calculated as the standard deviation (SD) of the loop length (l) measured in phosphodiester bonds. For multiple multi- or internal loops, A_{IML} is the mean SD of the loop length. ΔG -values were calculated using ViennaRNA 2.5.0 (42) using RNAfold for monomolecular constructs and RNAmultifold for bi-, tri- or tetra-molecular constructs. All ΔG -values were extrapolated to 27°C, the optimal growth temperature of insect-stage trypanosomes. The Bayesian information criterion (BIC) was calculated as $BIC = \ln(n) \times k - 2 \ln(\hat{L})$ with n = number of data points, k = number of parameters and \hat{L} = likelihood function.

RESULTS

Probing the solvent accessibility of *T. brucei* pre-mRNAs when bound to the editosome

Of the 12 edited transcripts in the mitochondria of *T. brucei*, we selected five pre-mRNAs for the analysis: the pre-mRNA of subunit 6 of the mitochondrial ATPase (A6), the transcript encoding subunit 3 of the cytochrome c oxidase (CO3), the pre-mRNAs of subunits 3 and 7 of the NADH dehydrogenase (ND3, ND7) and the transcript of ribosomal protein S12 (RPS12). All five RNAs represent extensively edited transcripts and cover a size range from 336nt (RPS12) to 1057nt (ND7) (Figure 1B). The different RNAs were synthesized by run-off *in vitro* transcription with purities $\geq 98\%$ (Figure 1C). Following synthesis, the RNAs were refolded and editosome : pre-mRNA complexes were formed at a 2-fold molar excess of editosomes to ascertain that every RNA-molecule is editosome-bound. The RNA-solvent accessibility was analyzed by hydroxyl radical probing (HRP) (22,23). Reactions were performed at single-hit conditions and sites of HR-induced RNA-backbone cleavage were identified as abortive cDNA-synthesis products. Cleavage intensities were normalized to a scale from 0 to 4 with a value of 1.0 representing the median cleavage intensity. Values ≤ 0.5 represent solvent inaccessible ribonucleotides (22,23).

In toto, we probed 5130 nt positions comparing the five pre-mRNAs in their free and editosome-bound states. As shown in Figure 2 all pre-mRNAs display complex solvent accessibility profiles both, as free RNAs and when bound to the editosome. The different HRP-traces are characterized by alternating regions of cleaved backbone positions above and below the median. When filtered with a moving average over a 3nt window, these sequences span be-

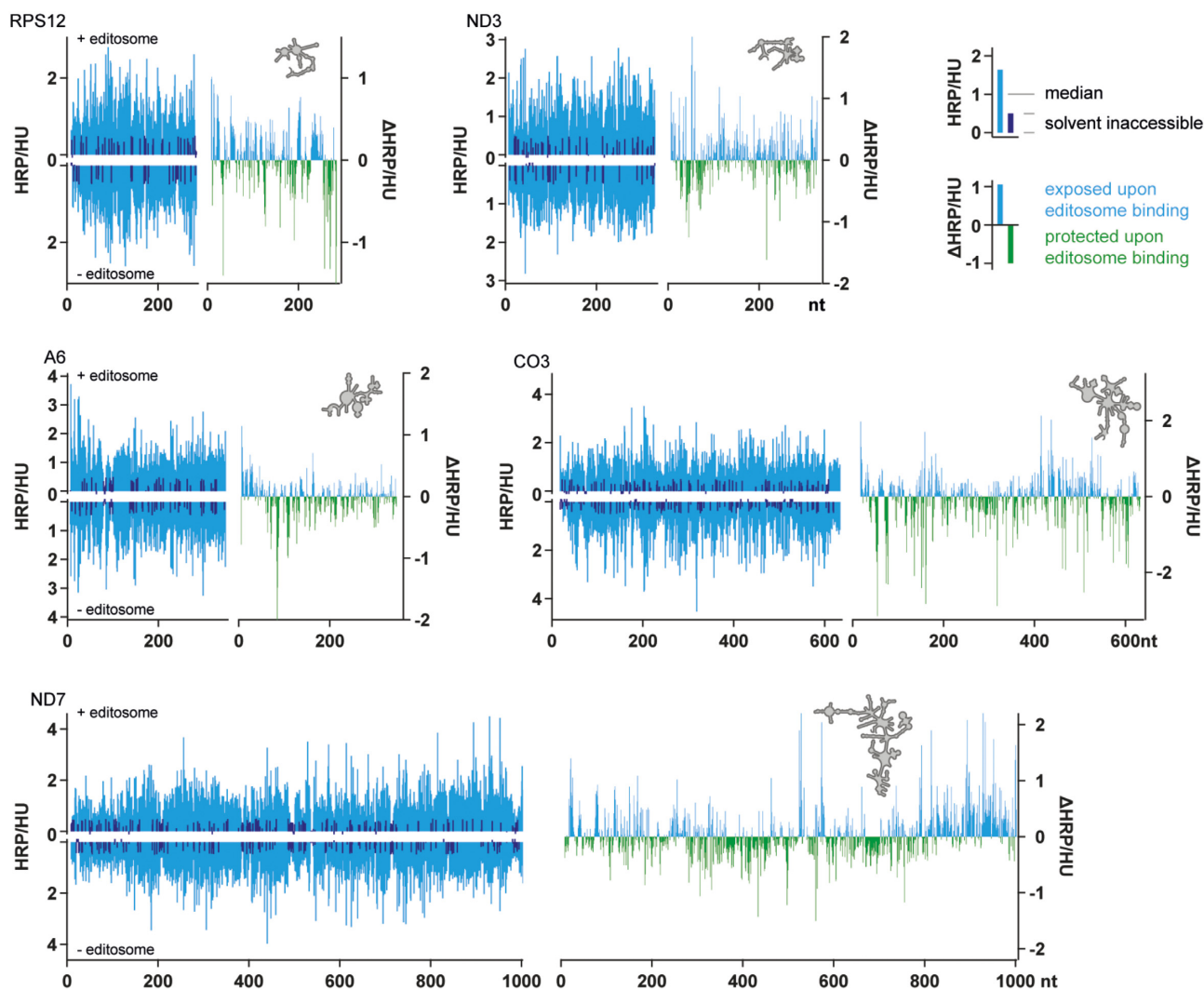


Figure 2. HRP-reactivity profiles. Left panels: Histograms of the mapped HRP-cleavage intensities (blue bars) versus nucleotide position for the RPS12-, ND3-, CO3-, A6- and ND7-pre-mRNAs in the presence (top) and absence (bottom) of editosomes. The data are normalized to the median reactivity and intensities ≤ 0.5 are considered solvent inaccessible (bars in dark blue). Right panels: Difference (Δ)HRP-plots of the cleavage intensities in the editosome-bound state minus cleavage intensities in the free RNA. Nucleotide positions that become protected upon editosome binding are in green and nucleotides with enhanced HRP-reactivities are in light blue. The 2D-structures of the different pre-mRNAs are shown as grey silhouettes. Replicates of the HRP-experiments correlate with Pearson correlation coefficients ($P \geq 0.9$). HU = HRP-unit.

tween 4–5nt in the free RNAs and 6nt in the editosome-bound states. Importantly, the HRP-profiles of the two folding states (free RNA versus editosome-bound RNA) differ for every pre-mRNA. This is reflected in a mean Pearson correlation coefficient of $P = 0.75$ between the two folding situations, which indicates that the solvent-accessibility of the 5 pre-mRNAs changes upon editosome binding. Localized differences can reach up to 3-times the median reactivity. To further analyze the data, we generated HRP-difference (Δ HRP)-plots by subtracting the nt-cleavage intensities of the free RNAs from the corresponding values of the editosome-bound state. This identifies solvent inaccessible nt-positions as negative values and nt-positions with increased solvent accessibility as positive values. Figure 2 shows the Δ HRP-plots for all 5 pre-mRNAs. In every case, solvent-protected nucleotides, as well as nucleotide positions with enhanced solvent accessibility, can be identified and are dispersed over the entire primary sequences. This

is consistent with the documented complex inherent RNA-chaperone activity of the editosome (11,15), which catalyzes a partial RNA-unfolding reaction to assist the formation of pre-mRNA/gRNA hybrid molecules to initiate the editing process (12). Furthermore, by plotting the number of highly protected nucleotides in the editosome-bound state as a function of all interrogated nucleotides in the different RNAs we identified a saturation-type of behavior (Supplementary Figure S5). A number of about 60nt are maximally involved in the pre-mRNA : editosome interaction, which suggests that the RNA-contact site of the editosome is a finite surface area.

Fuzzy RNA structure recognition by the editosome

Figure 3 shows the experimentally-derived 2D minimal-free-energy (MFE)-structures of all tested pre-mRNAs. Plotted on the different structures are the individual

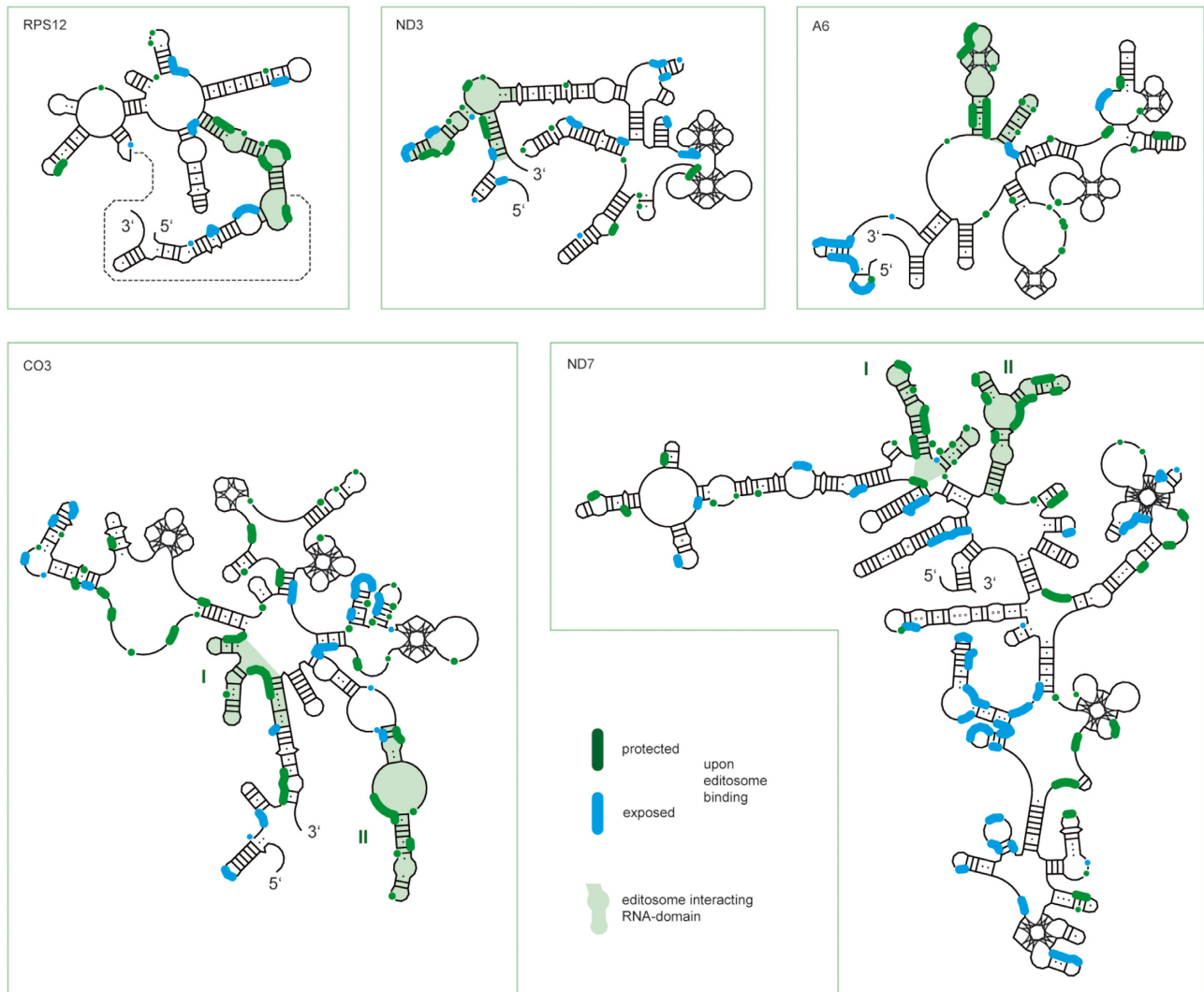


Figure 3. Secondary structure mapping of HRP-reactivities. Experimentally derived 2D structures of the *T. brucei* pre-mRNAs encoding RPS12, ND3, A6, CO3 and ND7 (14). Non-canonical and GU-bp are shown as dots. G-quadruplex elements are represented as leaf-like structures. A pseudoknot interaction in the RPS12 pre-mRNA is shown as a dotted line. Nucleotides protected from HRP-cleavage upon editosome binding (25th-percentile of negative $\Sigma\Delta\text{HRP}$ -values $\times 3$) are marked in green and nucleotides with increased solvent accessibility upon editosome binding (75th-percentile of positive $\Sigma\Delta\text{HRP}$ -values $\times 3$) are in blue. For the pre-mRNAs of ND7 and CO3, two editosome-interacting motifs were identified, numbered I and II.

ΔHRP -values of nucleotide positions that become protected from HR-cleavage upon editosome binding in green and nucleotides with increased solvent accessibility upon complex formation in blue. The data are suggestive of clustering of all affected nucleotides on the 2D-structure level, especially for the majority of solvent inaccessible nucleotides. For the two large pre-mRNAs ND7 (1057nt) and CO3 (705nt), two protected 2D-RNA elements can be distinguished. For the 336nt RPS12-transcript and the A6 (398nt) and ND3 pre-mRNAs (376nt), single motifs are discernable. The individual RNA domains have a length between 40–60nt. They differ in their base-pairing pattern and are characterized by a mean thermodynamic stability (ΔG) of -0.45 kcal/mol/nt and a conformational entropy of 0.55. Almost all of them branch off from multiloop arrangements and consist of irregular hairpins that contain next to multiple (38%) G:U bp, asymmetric loops, and bulged-out nucleotides. All attempts to identify a shared primary or secondary structure in the protected sequences failed. No in-

dividual nucleotide, nor a specific nucleotide sequence or 2D-motif is over or underrepresented in any of the footprinted sequences (Figure 4). However, the footprints are highly reproducible. Multiple experiments with different pre-mRNA preparations (≥ 5) and different editosome isolates (≥ 5) compare with a mean Pearson correlation coefficient of $P = 0.96$ proving that the identified footprints are not stochastically driven. Instead, they seem to indicate a high degree of conformational plasticity or in other words ‘fuzziness’.

Graph theory-based RNA-motif analysis

To quantitatively assess the topological characteristics of the different editosome-interacting RNA motifs, we scrutinized the protected RNA sequences in a coarse-grained analysis. For that, we used a graph theory-based approach as established by Schlick and colleagues (25,37,43,44). The method transforms the identified RNA 2D-elements (Fig-

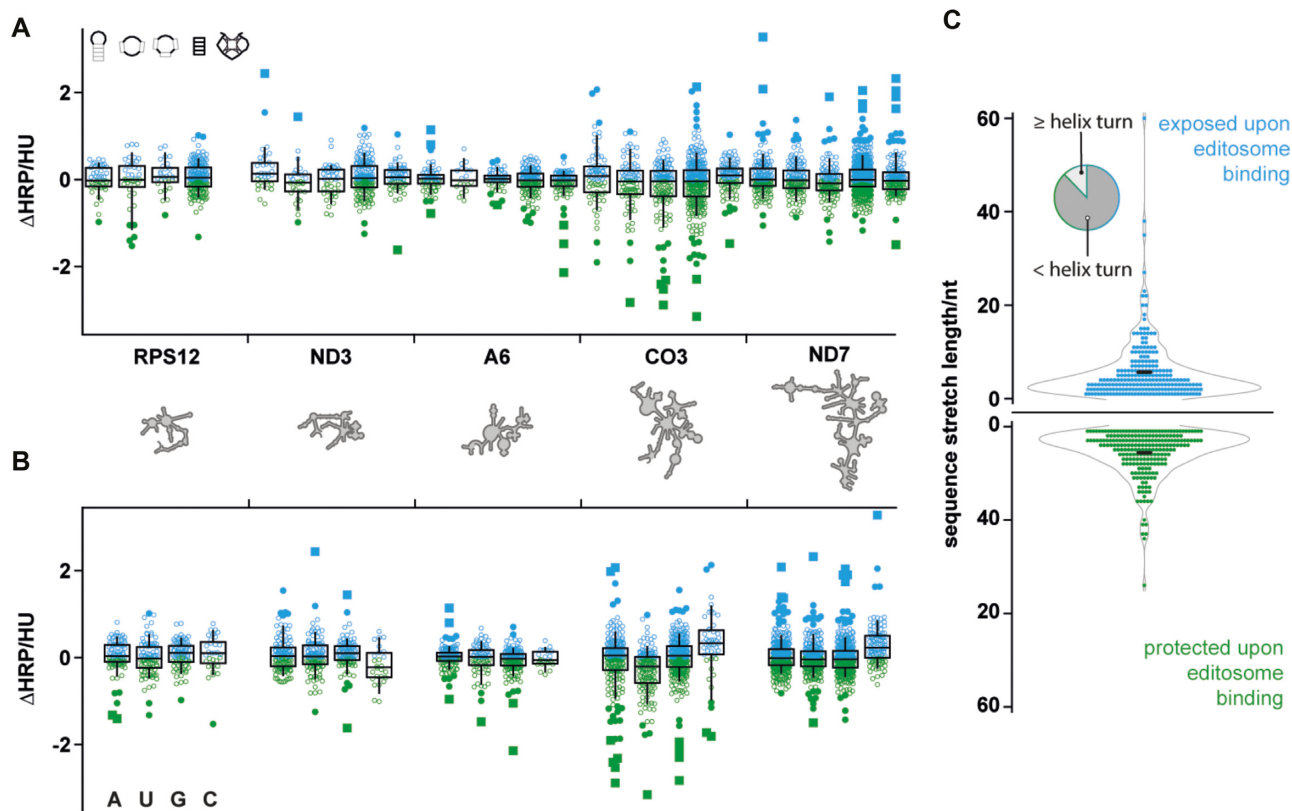


Figure 4. Statistical analysis of HRP-reactivity profiles. (A) Boxplot analysis of the Δ HRP-data for nucleotides in hairpin-loops, internal-loops, multi-loops, helices, and GQ-elements (from left to right) for all tested pre-mRNAs (RPS12, ND3, A6, CO3, ND7). Sketches of the different 2D-elements are depicted in the upper left corner. Blue: increased solvent accessibility upon editosome binding. Green: Protection upon editosome binding. Whiskers indicate the 9th and 91st percentile. Outliers are shown as filled circles ($1.5 \times$ interquartile range (IQR)) or as filled squares ($3 \times$ the IQR). The median is represented by a horizontal line. HU = HRP-unit. (B) The same plot as in (A) analyzing the nucleotide identity (A, U, G, C from left to right). (C) Length distribution of continuous sequence stretches with increased (blue) or decreased (green) solvent accessibility upon editosome binding after smoothing over a 3nt window. Solvent-exposed and solvent-protected sequence stretches are on average 6nt long (black horizontal bar), which is equal to half a helical turn. Only 13% of the sequence stretches are ≥ 11 nt (upper left pie chart). Further statistical data are provided in Supplementary Figures S5 and S6.

ure 5A) into discrete mathematical graph objects by describing the 2D-motifs as a set of vertices (V = loops, bulges, junctions) and edges (E = helices) (Figure 5B). This shrinks the complexity of the identified RNA elements down to their skeletal connectivity features and enables a quantitative comparison of the different RNA elements using spectral graph theory methods (37,45). The matrix of a graph specifies the degree of connectivity between the individual vertices. As suggested by Gan *et al.* (37), we used the Laplacian $V \times V$ -matrix (L), which was constructed from the adjacency (A) and square diagonal (D) matrices of the graphs as $L = D - A$. A specifies the number of edges that connect pairs of vertices and D details the connectivity of each vertex (Figure 5C). Furthermore, a V -vertex graph is characterized by its eigenvalue spectrum ($0 \leq \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \lambda_4 \dots$). This can be used to calculate graph similarities. Typically, RNA graphs are characterized by $\lambda_1 = 0$ and $\lambda_2 > 0$ (37). Thus, λ_2 reveals the overall connectivity of a graph, and graphs with similar λ_2 -values have similar topologies (Figure 5C). Lastly, we concatenated the individual adjacency (A)- and square diagonal (D)-matrices to generate 3D A - and D -matrix stacks. The stacks were averaged along the third-dimension to calculate mean A - and D -matrices (Figure 5D). The resulting averaged L -matrix

embodies a $4V(3E)$ -graph with an eigenvalue of 0.5477 (λ_2). Examples of two $4V(3E)$ -tree graphs are shown in Figure 5D.

Synthesis of RNA-motif library

To systematically unravel the complexity of the editosome-interacting $4V(3E)$ graph, we synthesized a library of topological RNA motifs. The library is centered on the $4V(3E)$ topology and extends the folding space to higher and lower graph complexities. Altogether, we synthesized 55 RNA constructs covering a tree graph vertex space from $0V(0E)$ to $5V(4E)$ (Figure 6A). The different RNAs are characterized by a wide range of molecular descriptors such as nt-length and composition, Gibbs free energy (ΔG), R/Y -ratio, and molecularity (summarized in Supplementary Table S2). The library also samples the contribution of symmetric and asymmetric bulges, of looped-out nucleotides and hairpin loops, and includes linear 2D-topologies as well as branched geometries such as 3- and 4-helix-junctions. All RNA constructs were assembled from chemically synthesized oligoribonucleotides (Supplementary Table S1) and were characterized by their electrophoretic mobility in native polyacrylamide gels (Supplementary Figure S4). To fur-

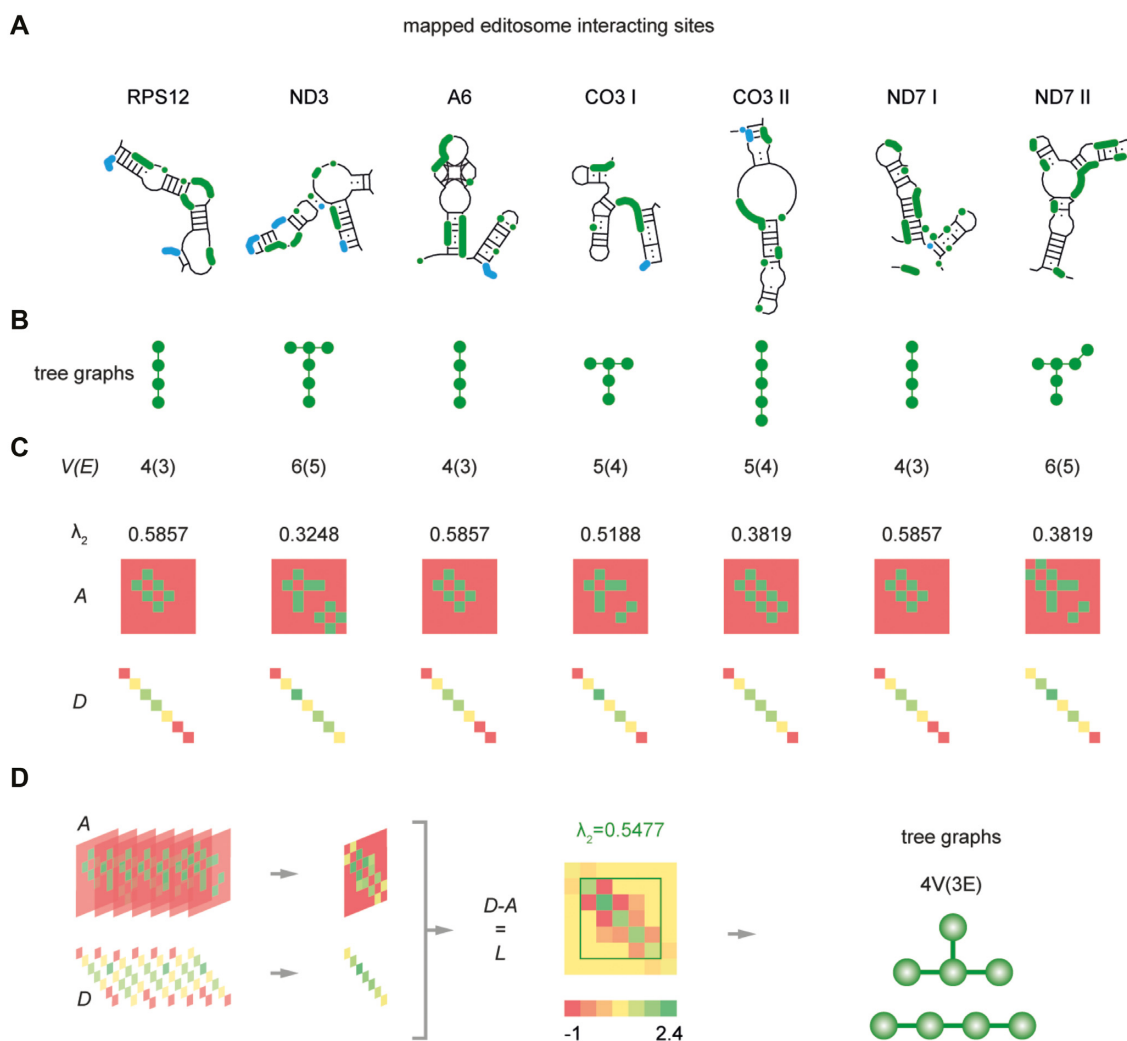


Figure 5. Graph theory-based RNA-motif analysis. (A) 2D-structures of the seven identified editosome interacting RNA motifs in the pre-mRNAs of RPS12, ND3, A6, CO3, and ND7. (B) Tree graph representations of the different RNA motifs. Vertices (V) = green dots. Edges (E) = connecting lines. (C) Matrix representations of the different RNA motifs: A = adjacency matrix, specifying the number of edges that connect pairs of graph vertices. D = square diagonal matrix detailing the connectivity of each vertex. λ_2 = second Laplacian eigenvalue, enumerating the overall pattern of connectivity. (D) 3D-matrix stacks are generated by concatenation of the individual A - and D -matrices. The stacks can be averaged along the 3rd-dimension to calculate mean A - and D -matrices. L = Laplacian matrix (L) defined as $L = D - A$. The resulting ‘averaged tree graph’ is $4V(3E)$ with a λ_2 eigenvalue of 0.5477. Two representative $4V(3E)$ -topologies are shown in green.

ther enumerate the topological diversity of the library, we applied next to the tree graph concept a second graph enumeration method known as dual graphs (37). Dual graphs invert the definition of the tree graph description: a vertex (V) represents a stem instead of a bulge, loop, or junction, and an edge (E) represents a single-stranded sequence instead of a stem. A side-by-side comparison of the tree- and dual-graph concepts is detailed in Figure 6B. Lastly, we calculated a set of topological indices for every RNA. The different descriptors represent established numerical invariants derived from spectral graph theory and chemical graph theory (46). They have been used in a large number of molecular topology studies (47–49) including pattern recognition studies in RNA (50). They include the Frobenius norm (FN), the Wiener index (WI), and the distance matrix energy (DE). All index calculations are detailed in the Material and Methods section. To enumerate the con-

nectivity of the different RNA constructs, we calculated relative connectivities (rC). rC values are rooted in the dual-graph concept and are based on the Randić index (51,52). Ignoring self-adjacency, rC values represent the ratio of the maximal trace of the degree matrix ($trace_{D_{max}}$) over the actual trace ($trace_D$). Altogether 23 descriptors were sampled for every RNA motif, which are summarized in Supplementary Table S2.

Editosome binding of the RNA-motif library

The aptitude of the different RNAs to bind to the editosome was analyzed by measuring the ability to saturate the single editosome RNA binding site (11) and as a result to block the processing reaction (Supplementary Figure S7). All measurements were performed in high-throughput in a concentration-dependent fashion, generating > 500 data

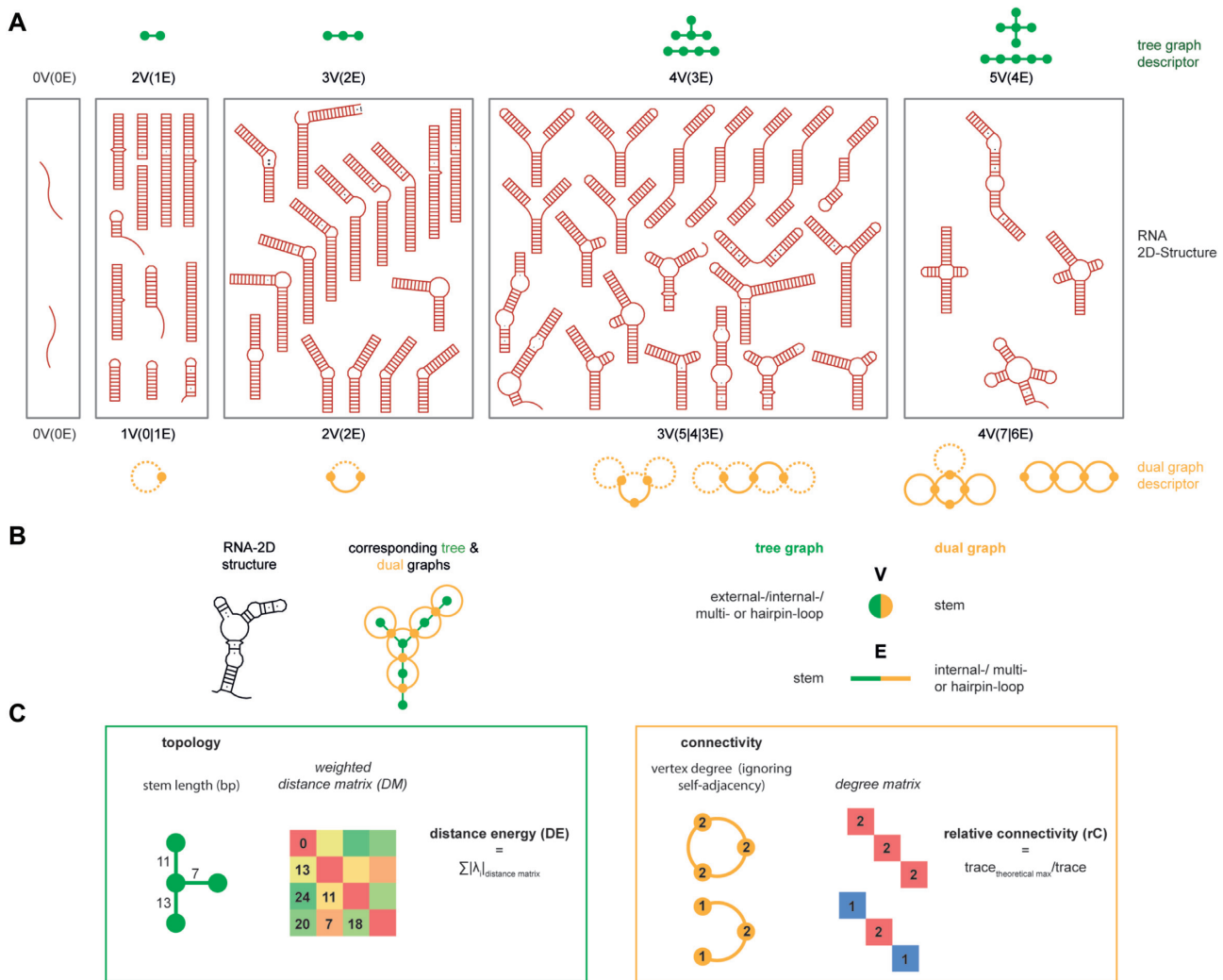


Figure 6. Topological characteristics of the RNA motif library. (A) 2D-structures of 55 RNA constructs varying in tree graph topologies between $0V(0E)$ and $5V(4E)$ (left to right). Tree graph annotations are shown on top of the 2D-structures in green and dual graph sketches are shown below in orange. (B) Left panel: Comparison of the tree- and dual-graph concepts using an arbitrary RNA 2D-element as an example. Right panel: Vertex (V) and edge (E) definitions for the tree graph and dual graph concepts. (C) Left panel: Definition of the distance energy (DE) as a topology descriptor based on the tree graph annotation. The distance matrix (DM) of a weighted tree graph stores the lengths of all shortest paths between every pair of vertices. DE represents the sum of all absolute eigenvalues ($|\lambda|$) of DM. Right panel: Definition of the connectivity concept based on the vertex degree of dual graphs, which is defined as the number of all incoming edges (ignoring the self-adjacency of hairpin loops). The relative connectivity (rC) calculates as the ratio of the maximal trace of the degree matrix ($\text{trace}_{D_{\max}}$) over the actual trace (trace_D). For dual graphs of the type $0V(0E)$ and $1V(0|1E)$ $rC = 0$. Edges not present in all dual graphs are shown as dotted lines.

points (Supplementary Figure S8). To quantitatively rank the different topologies, we calculated the molar excess (ME) required to inhibit the editing reaction halfmaximally (ME_{50}). As shown in Figure 7A, the collected ME_{50} -values spread over 4 orders of magnitude (from 0.9- to > 6000-fold). $0V(0E)$ and $2V(1E)$ tree-graphs invariably fail to interact with the editosome ($ME_{50} > 100$). Most higher-order graphs ($3V(2E)$ to $5V(4E)$) bind with low or only mediocre apparent strength ($10 < ME_{50} \leq 100$) and only ten RNA motifs bind with high affinity ($ME_{50} \leq 5$). As anticipated from the HRP-data, this group contains six $4V(3E)$ graphs and includes as the best editosome binder, a $4V(3E)$ 3-helix junction RNA with an ME_{50} of 0.9. Furthermore, three high-affinity interactors are $3V(2E)$ and one is $5V(4E)$. Collectively, this corroborates the above-described fuzziness of

the RNA motif (Figure 7B). Additionally, the data show that eight of the ten high-affinity interactors are of low connectivity ($rC > 1$). Increasing rC, for instance, by disrupting the phosphodiester backbone, invariably converts a mediocre or low-affinity binder ($10 < ME_{50} \leq 100$) into a decent or high-affinity interactor ($ME_{50} \leq 5$). Representative examples are shown in Figure 7C. This underscores the interplay between graph topology and graph connectivity and suggests a requirement for structural flexibility. To complete the analysis, we sampled all identified high-affinity graphs for the presence of a shared, isomorphic subgraph (53). As shown in Figure 8A, B, the $3V(2E)$ topology can be identified as the minimal building block in all $4V(3E)$ and $5V(4E)$ constructs. Depending on the specific $V(E)$ assignment it can be found in multiple orientations, which impli-

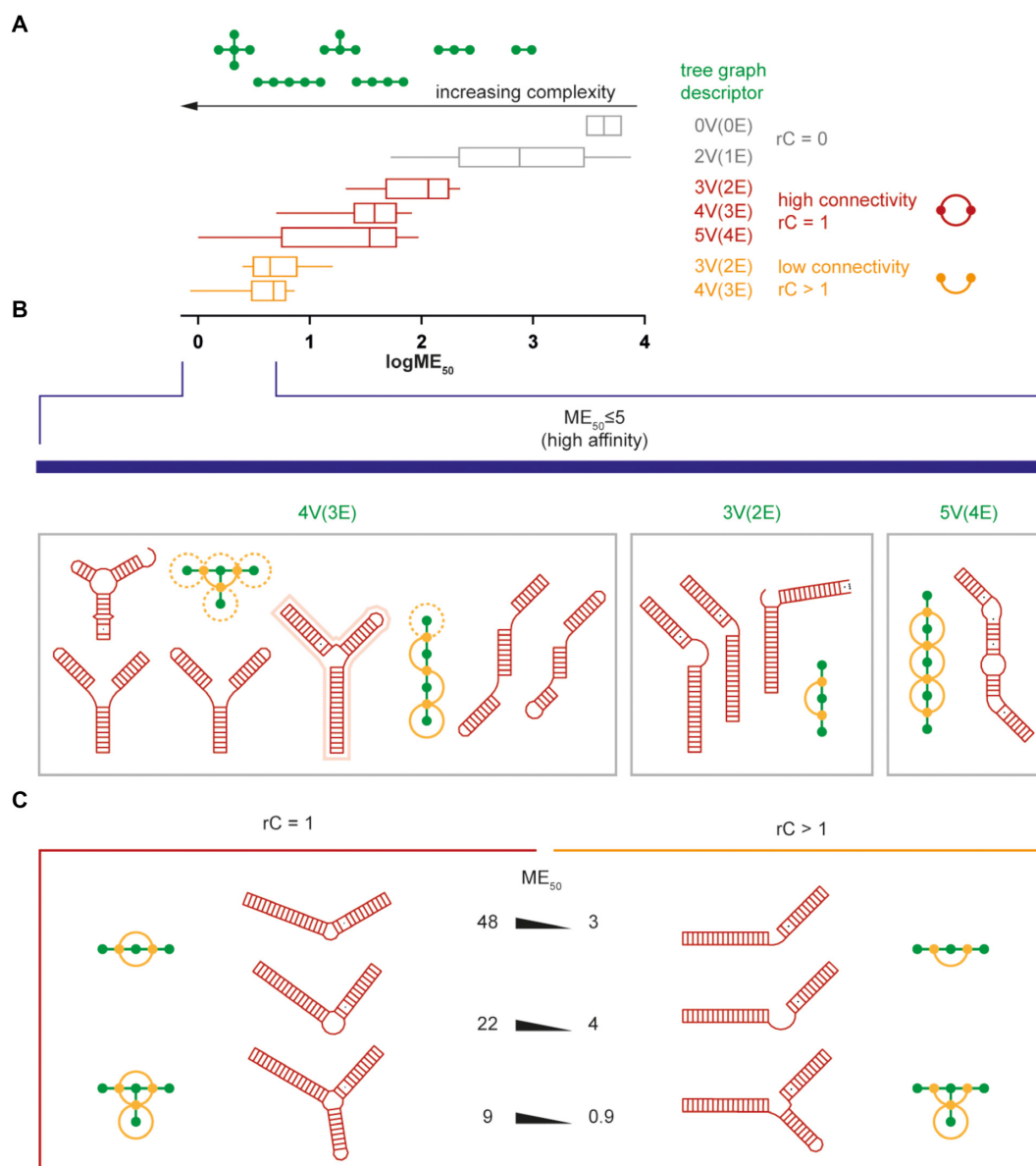


Figure 7. Editosome binding of the RNA motif library constructs. (A) Boxplot analysis of the measured ME₅₀-values of the different tree graph categories. Grey: 0V(0E) graphs. Red: High connectivity ($rC \leq 1$) RNA graphs (2V(1E), 3V(2E), 4V(3E), 5V(4E)). Orange: Low connectivity ($rC > 1$) 3V(2E) and 4V(3E) RNA constructs. Apparent editosome binding affinities are categorized as: ME₅₀ > 100 = no binding; 10 < ME₅₀ ≤ 100 = mediocre to low apparent binding strength; 5 < ME₅₀ ≤ 10 = good binding; ME₅₀ ≤ 5 = high apparent binding strength. (B) 2D-structures and graph annotations of all 4V(3E), 3V(2E), and 5V(4E) constructs with high editosome binding affinity (ME₅₀ ≤ 5). The best editosome interactor is a 4V(3E) ($rC = 1.5$) tree graph with ME₅₀ = 0.9 (outlined in light red). (C) Left: 2D-structures and graph annotations of two 3V(2E) and one 4V(3E) tree graph with relative connectivities of $rC \leq 1$. The three RNAs are characterized by ME₅₀-values of 48, 22 and 9 (top to bottom). Changing the connectivity by disrupting the phosphodiester backbone at single sites with rC values > 1 improves the apparent binding affinity by one order of magnitude.

icates that all higher graph topologies can be viewed as multiples of the minimal 3V(2E) motif. This likely indicates a hierarchical or modular setup of the editosome RNA binding motif and in this context, it is worth mentioning that the topology of the minimal gRNA/pre-mRNA substrates of the pre-cleaved *in vitro* RNA editing assay (54,16) is 3V(2E). Lastly, we harnessed the identified structural principles in a set of forward design experiments (Figure 8C). For that, we used several 2V(1E) motifs with no editosome binding activity (ME₅₀ > 100). However, when fused into single chimeric 4V(3E), $rC = 1.5$ topologies, the hybrid RNAs

as anticipated execute high-affinity binding characteristics (ME₅₀ ≤ 5).

Validation of the graph theory-based analysis

The results above suggest that both the graph topology and the relative connectivity (rC) represent key factors in modulating RNA binding to the editosome. However, both descriptors correlate or anticorrelate with other molecular features because they describe similar or co-dependent attributes (Figure 9A). For instance, the ΔG of an RNA

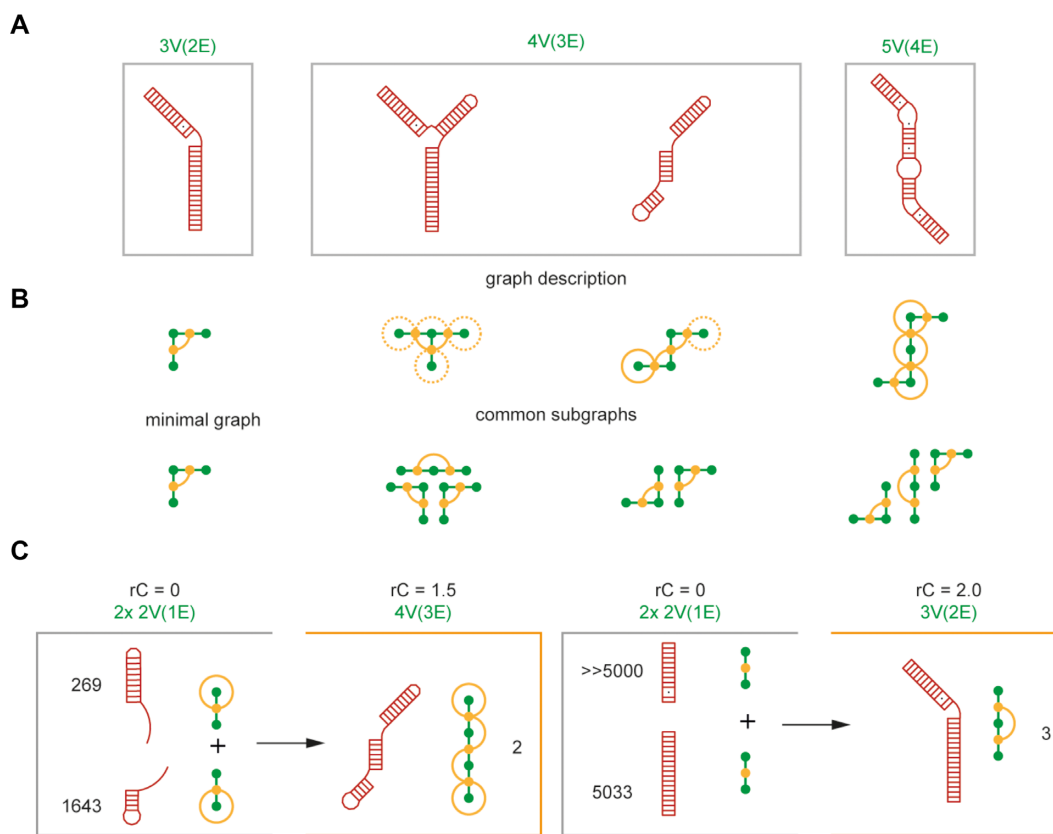


Figure 8. Subgraph analysis and forward design. (A) 2D-structures (red) of four representative high-affinity editosome interactors ($ME_{50} \leq 5$) with $3V(2E)$, $4V(3E)$, and $5V(4E)$ topologies. (B) Upper panel: Tree graph (green) and dual graph (orange) illustrations of the depicted RNAs. Lower panel: Subgraph analysis. The $3V(2E)$ topology can be identified in multiple orientations in all higher-order ($4V(3E)$ and $5V(4E)$) topologies. Thus, it represents the minimal subgraph of the editosome interaction RNA motif. (C) Forward design of two $3V(2E)$ high-affinity RNA graphs ($ME_{50} \leq 5$) from two $2V(1E)$ topologies that have no editosome binding affinity ($ME_{50} > 100$).

anticorrelates with the number of base-paired nucleotides and the $\Delta G/\text{nt}$ value correlates with the fraction of single-stranded nucleotides. Therefore, to pinpoint which of the 23 molecular descriptors of the RNA library predicts binding to the editosome best, we performed a multiple linear regression (MLR) analysis using ordinary least square fitting (OLS) involving all parameters. As shown in Figure 9B, the distance energy (DE) and the Frobenius norm of the distance matrix ($\|DM\|_{\text{FN}}$) outperform all other descriptors with a correlation coefficient (r^2) of 0.78 (BIC = 80.8). DE and rC together (as well as $\|DM\|_{\text{FN}}$ and rC together) result in r^2 -values of 0.84 (BIC = 66.9), which cannot be significantly improved by factoring in additional parameters. This demonstrates that about 85% of the editosome binding characteristics of an RNA motif can be predicted from just two descriptors: the distance energy (DE) or Frobenius norm of the distance matrix ($\|DM\|_{\text{FN}}$) and the relative connectivity (rC). To test this hypothesis, we performed an observed *versus* predicted (OP)-analysis. For that, we calculated ME_{50} -values for all 55 library RNA constructs using the two-factor DE/rC- or $\|DM\|_{\text{FN}}$ /rC-model and plotted the resulting values in relation to the measured ME_{50} -data. Figure 9C shows the corresponding scatter plot. Linear regression of the data points establishes a Pearson correlation coefficient of $P = 0.93$.

DISCUSSION

The U-nucleotide-specific RNA editing reaction in the mitochondria of kinetoplastid organisms such as African trypanosomes has been identified as an inherently noisy process. This is based on the observation that steady-state isolates of mitochondrial RNA contain next to pre-edited mRNAs large quantities of partially edited and mis-edited transcripts (17). As a consequence, editosomes, the protein complexes that catalyze the editing reaction, must be able to recognize a pool of RNA molecules that not only is highly divergent in primary sequence and nucleotide length but also in U-content. Moreover, both the primary sequences and the U-content constantly change as the reaction proceeds, which makes the binding of substrate RNAs into the single RNA-binding site of the editosome (11) a moving target problem. How editosomes untangle the structural multiplicity and dynamic disorder in the RNA-substrate landscape is unknown.

To unravel the RNA-recognition characteristics of the editosome, we determined the RNA-interaction sites of five *T. brucei* mitochondrial pre-mRNAs with nucleotide resolution. We used the differential RNA-solvent accessibility between the free and editosome-bound folding states as a read-out and probed the solvent accessibility by high throughput hydroxyl radical footprinting as established by

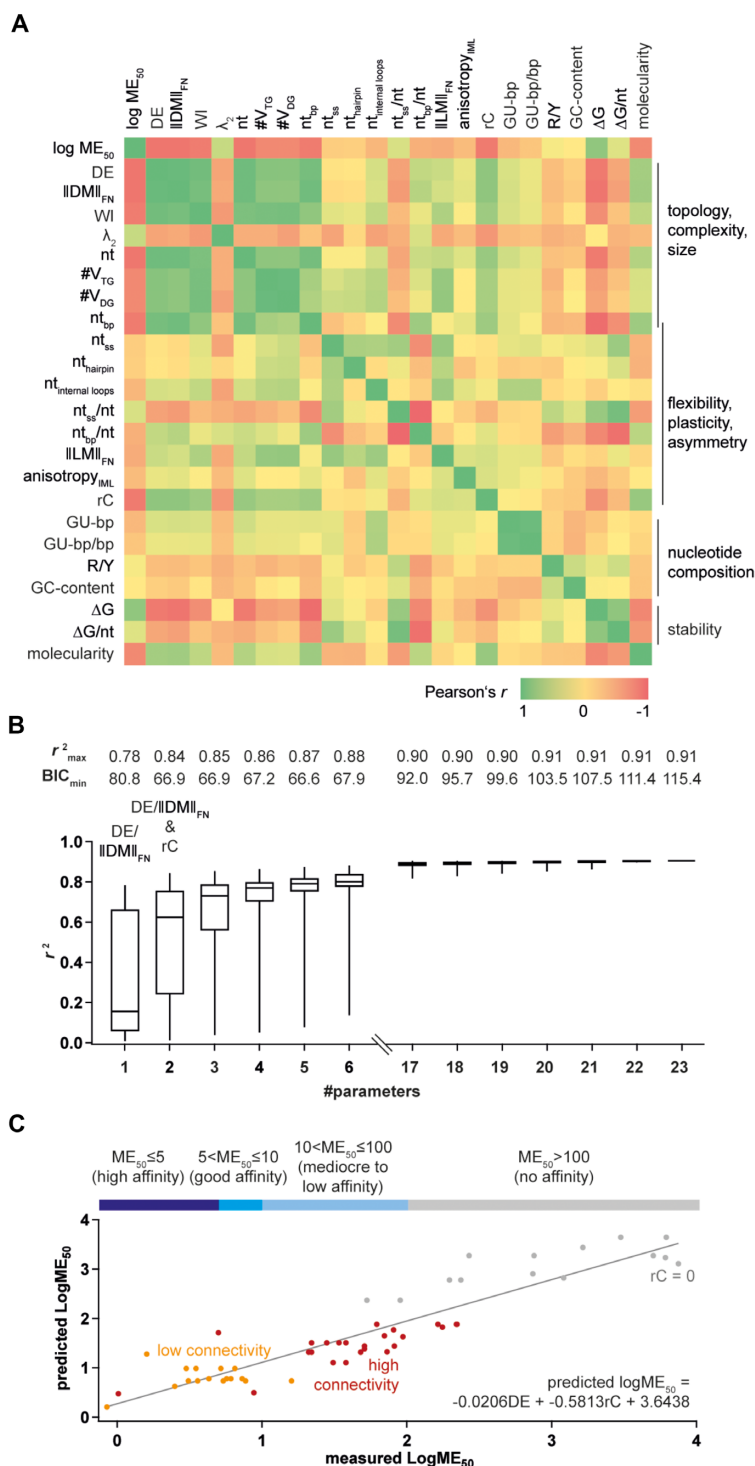


Figure 9. Multiple linear regression (MLR)-analysis of the relationship between the apparent editosome binding strength and the different RNA topology indices. **(A)** Heatmap matrix assessing the correlation of the measured ME_{50} -values to all 23 RNA motif descriptors. The set includes topology, complexity, and size descriptors including the distance energy (DE), the Frobenius norm of the distance matrix ($\|DM\|_{FN}$), and the Wiener index (WI). Other parameters indicate the flexibility, plasticity, and asymmetry of the RNA constructs as well as the relative connectivity (rC), nt composition, thermodynamic stability (ΔG), and molecularity. For a summary of all descriptors, see Supplementary Table S2. **(B)** MLR-analysis to scrutinize the dependence of the 23 descriptors to the $\log ME_{50}$ -values. The quality of the model is expressed in the form of r^2 -values for n over k models with $n = 23$ and $k = 1-6$ and $k = 18-23$. The calculated r^2 -values are shown as boxplots with whiskers representing minimal and maximal values. For all models, the maximal r^2 and the minimal Bayes Information Criterion (BIC) are listed. For one independent variable, DE and $\|DM\|_{FN}$ outperform all other descriptors with an r^2 of 0.78. This value can only be improved by adding rC as a second variable ($r^2 = 0.84$). Using more variables does not significantly improve r^2 but increases BIC. **(C)** Observed over predicted (OP)-plot of the $\log ME_{50}$ -values for all 55 RNA motifs using the DE/rC two-factor model. Data points are color-coded. Grey: $0V(0E)$ graphs. Red: High connectivity ($rC \leq 1$) RNA motifs $2V(1E)$, $3V(2E)$, $4V(3E)$ and $5V(4E)$. Orange: Low connectivity ($rC > 1$) RNA constructs $3V(2E)$ and $4V(3E)$. Linear regression results in a Pearson correlation coefficient of $P = 0.93$.

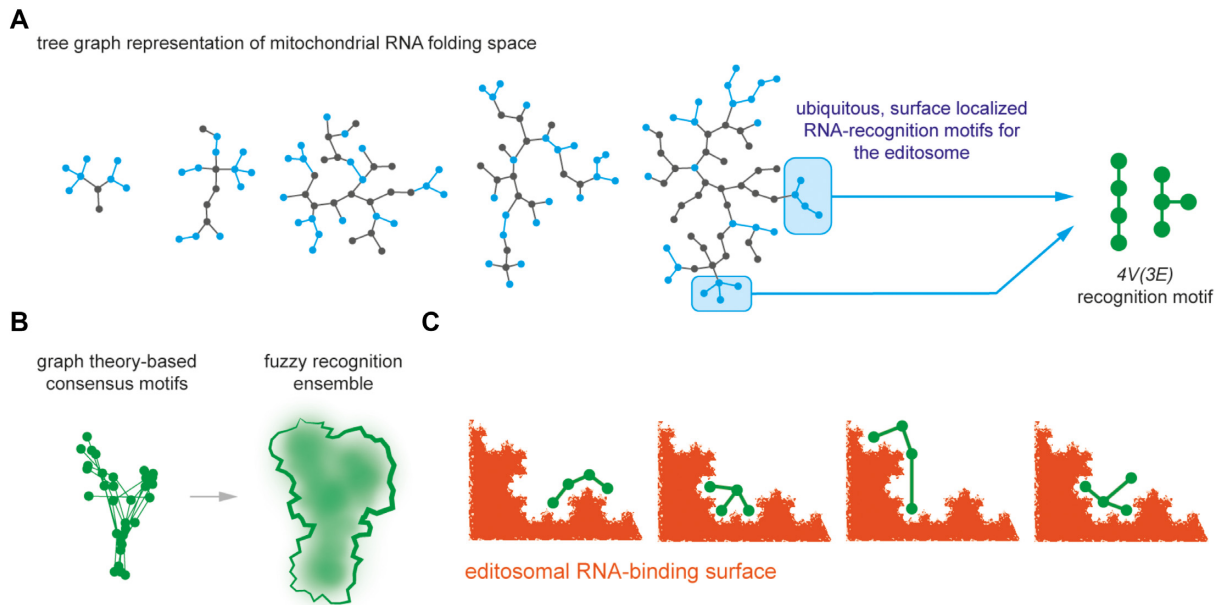


Figure 10. Fuzzy RNA recognition of the *T. brucei* editosome. (A) Tree graph representation of the convoluted RNA folding landscape of the *T. brucei* mitochondrial transcriptome. RNA species that are substrates of the RNA editing reaction are characterized by one or more editosome-binding motifs (highlighted in blue), which all fit the graph theory-derived 4V(3E)-criterion. (B) The multitude of similar but nonidentical motifs can be viewed as a fuzzy RNA ensemble, which enables the editosome to interact with a highly divergent RNA folding space. (C) We hypothesize that the RNA/editosome interface relies on multiple physicochemical interaction principles including strong and weak forces with a high degree of conformational entropy. This allows for structural flexibility and includes the capacity to self-modulate and self-adjust.

Weeks and colleagues (22,23). Editosomes have been shown to bind pre-edited mRNAs with nanomolar affinity (11) and thus, as anticipated, binding of the 800 kDa protein complex generated discernable footprints in all pre-mRNAs. The majority of solvent-protected nucleotides cluster in and around defined secondary structure elements in all transcripts, which vary in length between 40 and 60nt. In addition, we identified RNA-sequence stretches that become more solvent-exposed upon editosome binding. This demonstrates that the binding of editosomes induces a partial refolding of the different pre-mRNAs in line with the documented RNA-chaperone activity of the protein complex (11,12,15). Thus, as for many RNA/protein assemblies, the formation of a pre-mRNA/editosome complex involves an RNA motif recognition step and an RNA structure refolding step (18,19). However, unlike in other RNA/protein complexes, the identified editosome-interacting RNA motifs are similar but not identical. All attempts to identify a shared nucleotide sequence or common 2D-fold in the solvent-protected RNA sequences failed. This suggests that the shared characteristics of the identified RNA motifs are somehow concealed on the nucleotide level and as a consequence require a different analytical and conceptual framework. Since RNA secondary structures are essentially 2D-networks, they can be represented as 2D, coarse-grained graphs (25,43). By using only vertices and edges as descriptors we transformed all solvent-protected RNA motifs into planar, tree-like graph objects, which reduced the individual RNA motifs to their connectivity attributes. This in turn enabled a mathematical comparison of the graphs, ultimately permitting the description of a shared tree-graph consisting of 4 vertices and 3 edges (4V(3E)). Every iden-

tified solvent-protected RNA motif is a representative of this shared graph object, demonstrating that the RNA-recognition step of the editosome is characterized by a high degree of plasticity or ‘fuzziness’ even though it is of high affinity (55,56). The editosome does not recognize a precisely folded mRNA motif, instead, it allows for conformational ambiguity similar to what has been described for ‘fuzzy’ protein/protein complexes (24) or the immunomodulatory RNA binding protein Roquin and the RNA decay factors UPF1 and G3BP1 (57,58). The different proteins rely on RNA shape recognition rather than sequence recognition and as a consequence can process a large and diverse set of RNA ligands.

To pinpoint the defining topological constraints that drive the RNA/editosome interaction, we probed >50 topological RNA motifs ranging from 0V(0E) to 5V(4E). As before, the data confirmed the 4V(3E) motif as being optimal for a high-affinity interaction and identified the distance matrix energy (DE), the Frobenius norm of the distance matrix ($\|DM\|_{FN}$), and the relative connectivity (rC) as predictive indices for the binding performance. A combination of either DE and rC or $\|DM\|_{FN}$ and rC predict the propensity of any RNA motif to interact with the editosome with 85% certainty. This highlights the interplay between the general RNA topology and the connectivity of the individual RNA building blocks. Low connectivity values likely indicate a necessity for structural flexibility, which might be indicative of an induced fit type of scenario in the binding reaction (59). This might also explain the long-standing observation that pre-cleaved pre-mRNA substrates are much more efficiently edited *in vitro* than non-precleaved pre-mRNAs (54). The contribution of RNA

flexibility to the editosome binding reaction is further supported by the characteristics of the 5V(4E) construct. The RNA contains multiple symmetric and asymmetric bulges in addition to several G:U bp (Figure 7). This results in a highly flexible geometry (60), which as a consequence, displays high-affinity binding characteristics ($ME_{50} = 1$). In this context, it is important to note, that both, the 4V(3E) and 5V(4E) topologies can be deconstructed into multiple 3V(2E) subgraphs. This suggests a modular setup of the editosome RNA binding motif and rationalizes the topologies of the small RNA substrates of the standard pre-cleaved *in vitro* editing assays (16). The molecules consist of three short oligoribonucleotides that mimic two pre-mRNA fragments and a complementary gRNA. The resulting hybrid RNAs have 3V(2E), $rC > 1$ topologies, and as such they bind with high affinity to the editosome.

Importantly, none of the identified editosome-interacting RNA motifs in the different pre-mRNAs corresponds to a gRNA entry site or a prominent editing intermediate on the 2D level. Although we cannot exclude that these sites are proximal on the 3D level it demonstrates that editosomes can bind pre-mRNAs in the absence of gRNAs. This is supported by *in vitro* experiments, which show that a pre-assembly of the two RNAs is not required for correct editing. However, it is important to note that gRNA/pre-mRNA-hybrid RNAs cannot be excluded as editosome recognition elements. The hybrid RNAs adopt archetypical 4V(3E) three-way junction folds and as such, they likely bind with high affinity. However, whether the dynamic exchange of gRNAs all through the reaction cycle permits such a scenario is unclear at this stage.

Which of the editosomal protein components can execute a fuzzy RNA-binding functionality? While only 4 of the roughly 20 editosomal proteins contain canonical RNA-binding domains (13) recent RNA-interactome screening experiments in human cells revealed that many RNA-binding proteins (55%) lack archetypical RNA-binding domains. Instead, they frequently harbor intrinsically disordered protein (IDP)-domains (19). IDPs bind RNA with high specificity by providing large complementary binding surfaces, which depend on structural flexibility and a coupling of the folding and binding steps. Importantly, editosomal proteins are exceptionally rich in IDP-domains. Nearly 30% of the amino acid sequences are disordered, which is higher than the average of the human proteome (22%) (13). Particularly, the six oligonucleotide/oligosaccharide-binding (OB)-fold proteins and here specifically the editosomal core proteins KREPA1, KREPA2, and KREPA3 are significantly more disordered than the average of all other editosomal proteins. As demonstrated by Voigt et al., 2018 (12), five of the six OB-fold proteins execute RNA-chaperone activity to partially resolve the highly folded pre-edited substrate mRNAs (14,15). The activity is connected to the IDP-domains of the proteins, which have been proposed to be localized on the surface of the high-molecular-mass complex (12). Thus, we hypothesize that the demonstrated fuzzy RNA-binding property of the *T. brucei* editosome is executed by the IDP-domains of some or all editosomal OB-fold proteins (Supplementary Figure S9). This is supported by the fact that ID proteins have been shown to sample a broad range of dynamic structural changes upon

interacting with partner molecules. This includes the formation of fuzzy protein complexes, which are characterized by retaining a disordered folding state during complex formation (24). Prominent examples are the assembly of RNP-stress granules (61), protein liquid-liquid phase separation (62,63), and RNA/capsid-protein interactions in viral systems (64,65). We propose that the identified fuzzy RNA motifs in the *T. brucei* pre-mRNAs mirror the spatially disordered folding characteristics of the editosomal IDP-domains, which enable the protein complex to interact with multiple mitochondrial RNA species and multiple RNA-recognition motifs simultaneously. As a result, the editosome can address the highly divergent mitochondrial RNA-folding space including all pre-, partially and mis-edited mRNAs in addition to the large pool of gRNAs. Furthermore, we predict that the RNA-recognition step of the editosome follows a disorder-to-disorder transition in which both the protein and the RNA-contact sites remain in a state of high conformational entropy (66) (Figure 10). Such a setup is characterized by a high degree of structural flexibility and includes the capacity to self-modulate and self-adjust. By relying on multiple physicochemical interaction principles involving attractive and repulsive forces (67), a fuzzy molecular interaction is not only able to dynamically modulate the binding reaction, but it is also able to respond to environmental fluctuations (66,68–70). This phenomenon might be responsible for the documented ability of the *T. brucei* editosome to edit mitochondrial mRNAs in both parasite life cycle stages at dissimilar redox and temperature conditions (27°C in insect-stage trypanosomes and 37°C in bloodstream-stage parasites).

Lastly, we predict that many more nonbinary, i.e. fuzzy RNA/protein interactions exist. Especially in situations in which proteins or protein complexes are required to respond to changing cellular conditions by scanning large ensembles of different RNA molecules (56). As such, our finding represents a logical extension of the fuzzy protein theory for disordered proteins (69,70) by merging the fuzziness and frustration concepts in the energy landscape of proteins with that of RNA molecules (71). As much as frustration in fuzzy protein/protein complexes causes a multiplicity of specific interactions (72,73), it is conceivable that frustration in RNA/protein complexes is responsible for the observed nonbinary pre-mRNA-binding characteristics of the *T. brucei* editosome.

DATA AVAILABILITY

Experimental data and computer programs for performing all data analyses described herein are available from the corresponding author upon request.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Andreas Völker for trypanosome cell growth, Robert Knieß and Cristian Del Campo for editosomes preparations, and Reza Salavati, Inna Aphasizheva and Ruslan Aphasizhev for transgenic trypanosome strains.

FUNDING

German Research Foundation [DFG-GO516/8-1]; Dr Illing-Foundation for Molecular Chemistry (to H.U.G.). Funding for open access charge: Deutsche Forschungsgemeinschaft (DFG).

Conflict of interest statement. None declared.

REFERENCES

- Cruz-Reyes, J., Mooers, B.H.M., Doharey, P.K., Meehan, J. and Gulati, S. (2018) Dynamic RNA holo-editosomes with subcomplex variants: insights into the control of trypanosome editing. *Wiley Interdiscip. Rev. RNA*, **2**, e1502.
- Göringer, H.U. (2012) 'Gestalt,' composition and function of the *Trypanosoma brucei* editosome. *Annu. Rev. Microbiol.*, **66**, 65–82.
- Müller, U.F., Lambert, L. and Göringer, H.U. (2001) Annealing of RNA editing substrates facilitated by guide RNA-binding protein gBP21. *EMBO J.*, **20**, 1394–1404.
- Müller, U.F. and Göringer, H.U. (2002) Mechanism of the gBP21-mediated RNA/RNA annealing reaction: matchmaking and charge reduction. *Nucleic Acids Res.*, **30**, 447–455.
- Mehta, V., Moshiri, H., Srikanth, A., Kala, S., Lukeš, J. and Salavati, R. (2020) Sulfonated inhibitors of the RNA editing ligases validate the essential role of the MRP1/2 proteins in kinetoplastid RNA editing. *RNA*, **26**, 827–835.
- Missel, A., Souza, A.E., Nörskau, G. and Göringer, H.U. (1997) Disruption of a gene encoding a novel mitochondrial DEAD-box protein in *trypanosomabrucei* affects edited mRNAs. *Mol. Cell. Biol.*, **17**, 4895–4903.
- Li, F., Herrera, J., Zhou, S., Maslov, D.A. and Simpson, L. (2011) Trypanosome REH1 is an RNA helicase involved with the 3'-5' polarity of multiple gRNA-guided uridine insertion/deletion RNA editing. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 3542–3547.
- Kumar, V., Ivens, A., Goodall, Z., Meehan, J., Doharey, P.K., Hillhouse, A., Osorio Hurtado, D., Cai, J.J., Zhang, X., Schnauffer, A. et al. (2020) Site-specific and substrate-specific control of accurate mRNA editing by a helicase complex in trypanosomes. *RNA*, **26**, 1862–1881.
- Blum, B., Bakalara, N. and Simpson, L. (1990) A model for RNA editing in kinetoplastid mitochondria: "guide" RNA molecules transcribed from maxicircle DNA provide the edited information. *Cell*, **60**, 189–198.
- Koslowsky, D., Sun, Y., Hindenach, J., Theisen, T. and Lucas, J. (2014) The insect-phase gRNA transcriptome in *trypanosomabrucei*. *Nucleic Acids Res.*, **42**, 1873–1886.
- Böhm, C., Katari, V.S., Brecht, M. and Göringer, H.U. (2012) *Trypanosoma brucei* 20S editosomes have one RNA substrate-binding site and execute RNA unwinding activity. *J. Biol. Chem.*, **287**, 26268–26277.
- Voigt, C., Dobrychlop, M., Kruse, E., Czerwoniec, A., Kasprzak, J.M., Bytner, P., Campo, C.D., Leeder, W.M., Bujnicki, J.M. and Göringer, H.U. (2018) The OB-fold proteins of the *trypanosomabrucei* editosome execute RNA-chaperone activity. *Nucleic Acids Res.*, **46**, 10353–10367.
- Czerwoniec, A., Kasprzak, J.M., Bytner, P., Dobrychlop, M. and Bujnicki, J. (2015) M. Structure and intrinsic disorder of the proteins of the *trypanosomabrucei* editosome. *FEBS Lett.*, **589**, 2603–2610.
- Leeder, W.M., Hummel, N.F. and Göringer, H.U. (2016) Multiple G-quartet structures in pre-edited mRNAs suggest evolutionary driving force for RNA editing in trypanosomes. *Sci. Rep.*, **6**, 29810.
- Leeder, W.M., Voigt, C., Brecht, M. and Göringer, H.U. (2016) The RNA chaperone activity of the *Trypanosoma brucei* editosome raises the dynamic of bound pre-mRNAs. *Sci. Rep.*, **6**, 19309.
- Del Campo, C., Leeder, W.M., Reißig, P. and Göringer, H.U. (2020) Analyzing editosome function in high-throughput. *Nucleic Acids Res.*, **48**, e99.
- Zimmer, S.L., Simpson, R.M. and Read, L.K. (2018) High throughput sequencing revolution reveals conserved fundamentals of U-indel editing. *Wiley Interdiscip. Rev. RNA*, **11**, e1487.
- Jankowsky, E. and Harris, M.E. (2015) Specificity and nonspecificity in RNA-protein interactions. *Nat. Rev. Mol. Cell Biol.*, **16**, 533–544.
- Balcerak, A., Trebinska-Stryjewska, A., Konopinski, R., Wakula, M. and Grzybowska, E.A. (2019) RNA- protein interactions: disorder, moonlighting and junk contribute to eukaryotic complexity. *Open Biol.*, **9**, 190096.
- Costa, M. and Monachello, D. (2014) Probing RNA folding by hydroxyl radical footprinting. In: waldsich c. (ed) RNA folding. *Methods Mol. Biol.*, **1086**, 119–142.
- Balasubramanian, B., Pogozelski, W.K. and Tullius, T.D. (1998) DNA strand breaking by the hydroxyl radical is governed by the accessible surface areas of the hydrogen atoms of the DNA backbone. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 9738–9743.
- Caia, D.S., Duncan, C.D. and Weeks, K.M. (2010) Nonhierarchical ribonucleoprotein assembly suggests a strain-propagation model for protein-facilitated RNA folding. *Biochemistry*, **49**, 5418–5425.
- Caia, D.S., Duncan, C.D. and Weeks, K.M. (2010) The mrs1 splicing factor binds the b13 group i intron at each of two tetraloop-receptor motifs. *PLoS One*, **5**, e8983.
- Sharma, R., Raduly, Z., Miskei, M. and Fuxreiter, M. (2015) Fuzzy complexes: specific binding without complete folding. *FEBS Lett.*, **589**, 2533–2542.
- Schlick, T. (2018) Adventures with RNA graphs. *Methods*, **143**, 16–33.
- Brun, R. and Schönenberger, M. (1979) Cultivation and *in vitro* cloning of procyclic culture forms of *trypanosomabrucei* in a semi-defined medium. *Acta Trop.*, **36**, 289–292.
- Gerace, E. and Moazed, D. (2015) Affinity purification of protein complexes using TAP tags. *Methods Enzymol.*, **559**, 37–52.
- Golas, M.M., Böhm, C., Sander, B., Effenberger, K., Brecht, M., Stark, H. and Göringer, H.U. (2009) Snapshots of the RNA editing machine in trypanosomes captured at different assembly stages *in vivo*. *EMBO J.*, **28**, 766–778.
- Kala, S., Moshiri, H., Mehta, V., Yip, C.W. and Salavati, R. (2012) The oligonucleotide binding (OB)-fold domain of KREPA4 is essential for stable incorporation into editosomes. *PLoS One*, **7**, e46864.
- Ringpis, G.E., Aphasizheva, I., Wang, X., Huang, L., Lathrop, R.H., Hatfield, G.W. and Aphasizhev, R. (2010) Mechanism of u insertion RNA editing in trypanosome mitochondria: the bimodal TUTase activity of the core complex. *J. Mol. Biol.*, **399**, 680–695.
- Leeder, W.M., Kruse, E. and Göringer, H.U. (2021) Trypanosomatid, fluorescence-based *in vitro* U- insertion/U-deletion RNA-editing (FIDE). *Biol. Protoc.*, **11**, e3935.
- Katari, V.S., van Esdonk, L. and Göringer, H.U. (2013) Molecular crowding inhibits U-insertion/deletion RNA editing *in vitro*: consequences for the *in vivo* reaction. *PLoS One.*, **8**, e83796.
- Fenton, H.J.H. (1894) Oxidation of tartaric acid in presence of iron. *J. Chem. Soc. Trans.*, **65**, 899–911.
- Leeder, W.M. and Göringer, H.U. (2020) Mapping the RNA chaperone activity of the *t. brucei* editosome using SHAPE chemical probing. *Methods Mol. Biol.*, **2106**, 161–178.
- Vasa, S.M., Guex, N., Wilkinson, K.A., Weeks, K.M. and Giddings, M.C. (2008) ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA*, **14**, 1979–1990.
- Thiel, B.C., Beckmann, I.K., Kerpedjiev, P. and Hofacker, I.L. (2019) 3D based on 2D: calculating helix angles and stacking patterns using forgi 2.0, an RNA python library centered on secondary structure elements. *F1000Research*, **8**, 287.
- Gan, H.H., Fera, D., Zorn, J., Shiffeldrim, N., Tang, M., Laserson, U., Kim, N. and Schlick, T. (2004) RAG: RNA-As-Graphs database - concepts, analysis, and features. *Bioinformatics*, **20**, 1285–1291.
- Fera, D., Kim, N., Shiffeldrim, N., Zorn, J., Laserson, U., Gan, H.H. and Schlick, T. (2004) RAG: RNA-As- Graphs web resource. *BMC Bioinf.*, **5**, 88.
- Wiener, H.J. (1947) Structural determination of paraffin boiling points. *J. Am. Chem. Soc.*, **69**, 17–20.
- Galvez, J., Villar, V.M., Galvez-Llompert, M. and Amigó, J.M. (2011) Chemistry explained by topology: an alternative approach. *Comb. Chem. High Throughput Screen.*, **14**, 279–283.
- Ford, W. (2015) Vector and matrix norms. In: Ford, W. (ed) *Numerical Linear Algebra with Applications*. Academic Press, pp. 119–144.
- Lorenz, R., Bernhart, S.H., Höner zu Siederdisen, C., Tafer, H., Flamm, C., Stadler, P.F. and Hofacker, I.L. (2011) ViennaRNA package 2.0. *Algorith. Mol. Biol.*, **6**, 26.

43. Gan, H.H., Pasquali, S. and Schlick, T. (2003) Exploring the repertoire of RNA secondary motifs using graph theory; implications for RNA design. *Nucleic Acids Res.*, **31**, 2926–2943.
44. Izzo, J.A., Kim, N., Elmetwaly, S. and Schlick, T. (2011) RAG: an update to the RNA-As-Graphs resource. *BMC Bioinf.*, **12**, 219.
45. Cvetkovic, D.M., Doob, M. and Sachs, H. (1995) In: *Spectra of Graphs*. Johann Ambrosius Barth Verlag, Heidelberg.
46. Amigó, J.M., Gálvez, J. and Villar, V.M. (2009) A review on molecular topology: applying graph theory to drug discovery and design. *Naturwissenschaften*, **96**, 749–761.
47. Duarte, M.J., García-Domenech, R., Gálvez, J., Alemán, P.A., Martín-Algarra, R.V. and Antón-Fos, G.M. (2006) Application of a mathematical topological pattern of antihistaminic activity for the selection of new drug candidates and pharmacology assays. *J. Med. Chem.*, **49**, 3667–3673.
48. Dudek, A.Z., Arodz, T. and Gálvez, J. (2006). Computational methods in developing quantitative structure-activity relationships (QSAR): a review. *Comb. Chem. High Throughput Screen.*, **9**, 213–228.
49. Daianu, M., Mezher, A., Jahanshad, N., Hibar, D.P., Nir, T.M., Jack, C.R. Jr, Weiner, M.W., Bernstein, M.A. and Thompson, P.M. (2015) Spectral graph theory and graph energy metrics show evidence for the Alzheimer's disease disconnection syndrome in APOE-4 risk gene carriers. *Proc. IEEE Int. Symp. Biomed. Imaging*, **2015**, 458–461.
50. Benedetti, G. and Morosetti, S. (1996) A graph-topological approach to recognition of pattern and similarity in RNA secondary structures. *Biophys. Chem.*, **59**, 179–184.
51. Randić, M. (1975) On characterization of molecular branching. *J. Am. Chem. Soc.*, **97**, 6609–6615.
52. Gutman, I., Furtula, B. and Katanić, V. (2018) Randić index and information. *AKCE Int. J. Graphs Combinatorics*, **15**, 307–312.
53. Gawronski, A.R. and Turcotte, M. (2014) RiboFSM: frequent subgraph mining for the discovery of RNA structures and interactions. *BMC Bioinf.*, **15**(Suppl. 13), S2.
54. Carnes, J. and Stuart, K.D. (2007) Uridine insertion/deletion editing activities. *Methods Enzymol.*, **424**, 25–54.
55. Shen, Q., Shi, J., Zeng, D., Zhao, B., Li, P., Hwang, W. and Cho, J.H. (2018) Molecular mechanisms of tight binding through fuzzy interactions. *Biophys. J.*, **114**, 1313–1320.
56. Järvelin, A.I., Noerenberg, M., Davis, I. and Castello, A. (2016) The new (dis)order in RNA regulation. *Cell Commun. Signal*, **14**, 9.
57. Schlundt, A., Niessing, D., Heissmeyer, V. and Sattler, M. (2016). RNA recognition by roquin in posttranscriptional gene regulation. *Wiley Interdiscip. Rev. RNA*, **7**, 455–469.
58. Fischer, J.W., Busa, V.F., Shao, Y. and Leung, A.K.L. (2020) Structure-mediated RNA decay by UPF1 and G3BP1. *Mol. Cell.*, **78**, 70–84.
59. Williamson, J.R. (2000) Induced fit in RNA-protein recognition. *Nat. Struct. Biol.*, **7**, 834–837.
60. Mustoe, A.M., Bailor, M.H., Teixeira, R.M., Brooks, C.L. and Al-Hashimi, H.M. (2012). New insights into the fundamental role of topological constraints as a determinant of two-way junction conformation. *Nucleic Acids Res.*, **40**, 892–904.
61. Protter, D.S.W., Rao, B.S., Van Treeck, B., Lin, Y., Mizoue, L., Rosen, M.K. and Parker, R. (2018) Intrinsically disordered regions can contribute promiscuous interactions to RNP granule assembly. *Cell Rep.*, **22**, 1401–1412.
62. Wu, H. and Fuxreiter, M. (2016) The structure and dynamics of higher-order assemblies: amyloids, signalosomes, and granules. *Cell*, **165**, 1055–1066.
63. Boeynaems, S., Alberti, S., Fawzi, N.L., Mittag, T., Polymenidou, M., Rousseau, F., Schymkowitz, J., Shorter, J., Wolozin, B., Van Den Bosch, L. et al. (2018) Protein phase separation: a new phase in cell biology. *Trends Cell Biol.*, **28**, 420–435.
64. Ivanyi-Nagy, R., Lavergne, J.P., Gabus, C., Ficheux, D. and Darlix, J.L. (2008) RNA chaperoning and intrinsic disorder in the core proteins of flaviviridae. *Nucleic Acids Res.*, **36**, 712–725.
65. Byk, L.A. and Gamarnik, A. (2016) Properties and functions of the dengue virus capsid protein. *Annu. Rev. Virol.*, **3**, 263–281.
66. Miskei, M., Horvath, A., Vendruscolo, M. and Fuxreiter, M. (2020) Sequence-Based prediction of fuzzy protein interactions. *J. Mol. Biol.*, **432**, 2289–2303.
67. Gruet, A., Dosnon, M., Blocquel, D., Brunel, J., Gerlier, D., Das, R.K., Bonetti, D., Gianni, S., Fuxreiter, M., Longhi, S. et al. (2016) Fuzzy regions in an intrinsically disordered protein impair protein-protein interactions. *FEBS J.*, **283**, 576–594.
68. Horvath, A., Miskei, M., Ambrus, V., Vendruscolo, M. and Fuxreiter, M. (2020) Sequence-based prediction of protein binding mode landscapes. *PLoS Comput. Biol.*, **16**, e1007864.
69. Fuxreiter, M. (2020) Classifying the binding modes of disordered proteins. *Int. J. Mol. Sci.*, **21**, 8615.
70. Fuxreiter, M. (2020) Fuzzy protein theory for disordered proteins. *Biochem. Soc. Trans.*, **48**, 2557–2564.
71. Gianni, S., Freiberger, M.I., Jemth, P., Ferreira, D.U., Wolynes, P.G. and Fuxreiter, M. (2021) Fuzziness and frustration in the energy landscape of protein folding, function, and assembly. *Acc. Chem. Res.*, **54**, 1251–1259.
72. Freiberger, M.I., Wolynes, P.G., Ferreira, D.U. and Fuxreiter, M. (2021) Frustration in fuzzy protein complexes leads to interaction versatility. *J. Phys. Chem.*, **125**, 2513–2520.
73. Ferreira, D.U., Komives, E.A. and Wolynes, P.G. (2014) Frustration in biomolecules. *Q. Rev. Biophys.*, **47**, 285–363.